Data Engineering

By

W. N. Holmes, B.E.E., M.I.E.Aust.

Systems Development Institute, IBM Australia Ltd., Canberra

1 INTRODUCTION

Several circumstances conspire to make timely a proposal for formal establishment of a new branch of engineering.

The Institution of Engineers, Australia has published a scheme for separate colleges within the Institution for different branches of engineering (Storr 1973a). The increasing use of computing machinery involves an increasing number of engineers in development and use of computers. Courses of study in computing are being introduced to various levels of education. Professional engineering organisations are variously reacting to the influence of computing machinery - for example, the IEEE Computer Society has been formed from within the IEEE (McCluskey 1970), and the Institution of Engineers, Australia has been perturbed by "the increasing number of papers submitted for publication which describe the use of computers" (Harper 1970).

This paper proposes that a branch of engineering, perhaps best called "data engineering", be established by creating a "College of Data Engineering" in the Institution of Engineers, Australia, and that formal tertiary education in the discipline be sponsored by the Institution.

The scope and flavour of "data engineering" are illustrated in this paper, and course topics for tertiary study are tentatively sketched to give an idea of the possible sub-disciplines.

2 PROFESSIONAL USE OF COMPUTERS

(a) Computer-Based Occupations

Knowledge might be said to increase in breadth as we get to know about more and more phenomena, and in depth as we get to know more and more about particular phenomena. In engineering, and in other technical callings, increase of knowledge in depth brings with it the need to calculate more frequently and more precisely so that knowledge in depth can be exploited.

As a result, digital computers are increasingly used both to do calculations, and to manage calculations and their consequences. With the increase in numbers of digital computers, a priesthood has developed. A professional person wishing to make use of a computer often must apply to that priesthood to do so.

A miscellany of computer-based occupations has appeared, and the people filling those occupations have only recently had academic training to fit them specifically for that occupation. Now schools for programmers, and tertiary computer science courses (e.g. Spooner 1973), are springing up, and Societies of people interested in digital computers are striving to set up a formal profession.

(b) Spread of Digital Computers

The use of digital computers is becoming much more rapidly widespread than was anticipated, particularly since advanced manufacturing techniques such as "Large Scale Integration" have made it possible for the individual to buy electronic calculators from chemists' shops for less than $40, and for the widespread private use of digital computers, for example in cars, to be predicted as imminent.

Digital computers are becoming available and attractive to many more people than can ever be encompassed by a "computer science" profession. Indeed, there have long been many users of digital computers, skilled and demanding users, to whom their use of computers is an integral part of their practice of a profession unrelated directly to the theory of computation (e.g. Cutbill & Jones 1973).

Therefore, use of digital computers will become commonplace. Their use in schools as a basic item of study has been proposed (e.g. Arsac 1970). Their use in technical professions is already common, and bids fair to become a sine qua non in professional education (e.g. Committee on Physics in Two-Year Colleges 1973).

(c) A Computing Profession ?

However, just because computing machinery might come to be used by almost anyone, this does not mean that "computer scientists" or "computing practitioners" should similarly proliferate.

Teamwork between say, engineers and programmers (King 1973), has been espoused, but, when programming skills can be taught to primary school children (e.g. Papert 1970), the necessity must be questioned. It has even been questioned whether programming will survive as a calling (Willingham 1973).

In any case, the wisdom of founding a profession on a particular item of machinery must also be doubted (Holmes 1974).

It is suggested in this paper that a more fundamental profession, one in the general class of engineering, can be discerned, a profession neither based on nor limited by the use of digital computers.

3 DATA ENGINEERING

(a) Data as a Natural Resource

Data, according to an international standard vocabulary, are "a representation of facts or ideas in a formalised manner capable of being communicated or manipulated by some process." (Tootill 1968 p.3)

"Engineering is the art of applying the resources of nature, scientific principles and the accumulated experience of its practitioners for the use and convenience of mankind." (Storr 1972) If data are or come from a natural resource, then data engineering is a valid topic for discussion.

The assertion that life in a technological society is lived in a sea of data scarcely needs illustration. Rightly or wrongly, to be illiterate is to be severely handicapped.

That data are a resource is indicated by the organisations which traffic in data (e.g. Gruenberger 1968) much as electrical utilities traffic in electricity.

That data are a natural resource is shown by the instruments - radio telescopes, seismic arrays, weather satellites - deployed to collect data from nature.

Two objections may be put forward. Data are not always tangible. All data are not collected from nature. But compare data to electricity - electricity is not always tangible, nor is all electricity collected from nature, yet electrical engineering is an established branch of engineering.

(b) Data and Information

The two terms, "data" and "information", are often used as synonyms, but a useful, even vital, distinction should be made.

From the international standard vocabulary, information is "in automatic data processing the meaning that a human assigns to <u>data</u> by means of the known conventions <u>used</u> in its representation." (Tootill 1968 p.3) Thus, data are a representation, information is a meaning.

Appropriately, data engineering is concerned with data processes and data machines - for which data are, in turn, the raw material and the working fluid.

A major objective of data engineering would be to enhance the information conveyed by data, but information or meaning can not be manipulated by an engineered system except through the data which convey that information.

(c) The Name "Data Engineering"

Only minor practical problems are presented by the term "data engineering".

"Data" should be pronounced to rhyme with "freighter" (Shorter O.E.D.), though mispronunciation to rhyme with "barter" is common. It is awkward that the singular form of "data" is unpopular, and that "data" lacks a distinct adjectival form.

These grammatical problems are trivial. The real problem is the confusion of "data" with "information". If the standard distinction is not made between the two terms, then the study of the man-machine interface is gravely deficient in terminology, and that study is crucial to data engineering.

(d) Data Engineering versus Computer Science

The term "computer science" must be considered established, if only through its frequent use in print - as evidenced in the appended bibliography. The use of the term has been effectively supported (e.g. Newell et al. 1967), but its use in preference to say "data science" has not been argued.

Its weaknesses lie in its devotion to specific machinery and its acquisitiveness towards any calling displaying dependence on that machinery. Rapid changes in computer technology will force rapid changes in computer science. "Therein is the dilemma of computer science, being part mathematical science and part mathematical engineering." (Perlis 1968 p.71)

In contrast, data engineering is concerned with the manipulation and management of data. The internal workings of digital computers are of incidental interest only, and the craft of programming should be delegated to data engineering technicians (cf. Littler 1972). This is a far more stable basis for a profession.

(e) The Status of Data Engineering

Data engineering has no status of itself, as it has not been defined in the literature. Information engineering has been discussed (e.g. Slamecka 1968), computer engineering has been suggested (Kandel 1972), and software engineering has been written up (e.g. Buxton & Randall 1970, Tou 1970). Examples of data engineering have existed for some time (e.g. Sackman 1967).

Data engineering might be considered established through certain narrow technical disciplines - for example, management science & industrial engineering (Ackoff 1973) - themselves candidates for inclusion under data engineering.

4 EDUCATION IN DATA ENGINEERING

(a) Issues in Education

In the following description relating to education in data engineering, only generalities can be broached.

Details on present engineering education can be easily obtained (e.g. from Lloyd 1959), and computer science curricula of one kind or another abound (see the appended bibliography). From such sources and from the general suggestions below, a detailed data engineering education plan can be drawn up.

However, two opposing influences must be borne in mind - the influence toward increased technical content (Storr 1973b), and the influence toward increased generality and "relevance" (Davenport & Rosenthal 1967, Holliday 1966, Pickering 1973) even unto disestablishment of professions (Illich 1973).

(b) Data Engineering in the Schools

Technical education is not confined to tertiary institutions. It should begin when schooling begins. Therefore, any proposal for innovation in technical education should begin with a consideration of innovation in schools.

Papert (1970) believes that "the fundamental ingredients of educational innovation must be better things to do and better ways to think about oneself doing these things." If this belief is acted on, and if one accepts that "the essence of engineering is simply <u>problem solving under constraint</u>" (Scott 1973), then essential engineering must be introduced to schools. The question then is, what form should this essential engineering take ?

Children in schools are taught problem solving by example, but the principles of problem solving, essential engineering, are not taught explicitly. Arsac (1970) sees the best vehicle for this teaching to be "informatics", a science perhaps best defined (following Zemanek 1972) as having four components - formal description, process organisation, programming, and computer application.

(b) Data Engineering as a Basic Study

"Basically the engineer is concerned with three principal fields:
(1) management of information
(2) processing of materials
(3) conversion of energy" (Jansson 1973).

To support the first field, a general study of data engineering should be part of all basic engineering education.

By introducing a course in data engineering into the first year of engineering studies, valuable emphasis would be given to problem solving skills (Scott 1973), and the organisational and computational aspects in the study of other branches of engineering would be simplified and buttressed.

As things are now, students would necessarily be required to develop skill in programming, but not for the sake of the skill itself. The acquisition of programming skill might partly supplant the acquisition of draughting skill, particularly when digitally controlled draughting machines become cheaper.

(c) Data Engineering as a Principal Study

Following Slamecka (1968), one might carve data engineering into three parts - the use of data, data processes, and data systems.

Simply, the disciplines would study <u>data itself</u> (data statics? - representation, coding, recording, aggregation, modelling, grammar), the <u>manipulation of data</u> (data dynamics? - transmission, transformation, reduction, correction, simulation, algorithms, retrieval, parsing), and <u>the machines</u> for the manipulation of data (data mechanics? - systems, networks, switching, process control, logical design).

(d) Foundation Studies for Data Engineering

There would be two aspects to foundation studies for data engineering.

The first aspect is the modification of existing foundation studies to emphasise digital phenomena. For example, in mathematics this might mean stressing numerical analysis and logic, and a redirection of calculus from infinitesimals to finite differences.

The second aspect is the introduction of new fields of study - for example graph theory, information theory, and automata theory - to educate in concepts important to data engineering. Such new fields of study might be grouped under the heading of "informatics" to build on any studies introduced into the schools. The temptation to merely strip existing and proposed computer science courses of their engineering content should be resisted to allow a coherent course to be designed.

5 RECOMMENDATIONS

This paper has proposed recognition of a branch of engineering called "data engineering". The proposal has necessarily been in outline, not in detail - but an extensive bibliography is appended to allow interested readers to study particular issues.

Those who agree with the proposals should purposefully organise discussions with a view to formally recommending to the Institution whether
(1) tertiary education in data engineering should be sponsored by the Institution,
(2) a College of Data Engineering (with its Transactions) should be provided within the Institution (as well as, or instead of, the College of General Engineering).
(3) extensive primary and secondary education in "informatics" should be recommended to the authorities by the Institution acting in concert with other professional bodies.

6 BIBLIOGRAPHY

ACKOFF, RUSSELL L. (1973) Science in the Systems Age: Beyond IE, OR, and MS Operations Research 21, 3, MayJun73, 661-71.

ALT, FRANZ L; RUBINOFF, MORRIS; FREIBERGER, WALTER (Eds.1970) Advances in Computers - Volume 10 New York: Academic Press, 313pp.

AMERAL, SAUL (1971) Computer Science: A Conceptual Framework for Curriculum Planning Communications of the ACM 14, 6, Jun71, 391-401.

ARSAC, JACQUES J. (1970) Informatics and Computer Education, in Scheepmaker & Zinn, I/69-72.

ASHENHURST, R. L. (Ed.1972) A Report of the ACM Curriculum Committee on Computer Education for Management Communications of the ACM 15, 5, May72, 363-98.

ATCHISON, WILLIAM F. (Chairman 1968) Curriculum 68: Recommendations for Academic Programs in Computer Science Communications of the ACM 11, 3, Mar68, 151-97.

AUSTING ,RICHARD H; ENGEL, GERALD L. (1973) A Computer Science Course Program for Small Colleges Communications of the ACM 16, 3, Mar73, 139-47 or Management Informatics 1, 5, Oct72, 181-9.

BUXTON, JOHN; RANDELL, BRIAN (Eds.1970) Software Engineering Techniques NATO Science Committee, Apr70, 164pp.

COMMITTEE ON PHYSICS IN TWO-YEAR COLLEGES (1973) Computers in Physics Teaching American Journal of Physics 41, 10, Oct73, 1209-10 ("Physics faculty members in two-year colleges as well as those in four-year colleges and universities have a professional obligation to: (a) become proficient in the application of computers in teaching their own discipline; (b) ... ; (c) use computing services in their teaching whenever appropriate." from a position statement approved by unanimous ballot).

CUTBILL, JOHN L; JONES, KAREN SPARCK (1973) The Needs of the Pseudo-Pro Software - Practice and Experience 3, 2, AprJun73, 83-6.

DAVENPORT, WILLIAM H; ROSENTHAL, DANIEL (Eds.1967) Engineering: Its Role and Function in Human Society New York: Pergamon Press, Inc., 284pp. (an anthology presenting a selection of readings bridging the gap between the engineer and the humanist).

FINERMAN, AARON (Ed.1968) University Education in Computing Science New York: Academic Press, 237pp.

GRUENBERGER, FRED (Ed.1968) Computers and Communications - Toward a Computer Utility New Jersey: Prentice-Hall, Inc., 219pp.

HAMBLEN, JOHN W. (1971) Using Computers in Higher Education: Past Recommendations, Status, and Needs Communications of the ACM 14, 11, Nov71, 709-12.

HARPER, C.H.D. (Ed.1970) Papers Describing Computer Techniques Electrical Engineering Transactions, The Institution of Engineers, Australia, EE6, 1, Mar70, 45.

HOLLIDAY, LESLIE (Ed.1966) The Integration of Technologies London: Hutchinson & Co. Ltd., 167pp.

HOLMES, W. N. (1974) The Social Implications of the Australian Computer Society (submitted to) The Australian Computer Journal.

HOROWITZ, E; MORGAN, H.L.; SHAW, A.C. (1972) Computers and Society: A Proposed Course for Computer Scientists Communications of the ACM 15, 4, Apr72, 257-61.

ILLICH, IVAN (1973) The professions as a form of imperialism New Society 25, 571, 13Sep73, 633-5 (from "Tools for Conviviality" Calder & Boyars, 1973, 110pp.).

JANSSON, JAN-ERIK (1973) Tendencies in Engineering Education The Journal of the Institution of Engineers, Australia 45, 10-11, OctNov73, 10-12.

KANDEL, ABRAHAM (1972) Computer Science - A Vicious Circle Communications of the ACM 15, 6, Jun72, 470-1 (see also discussion op. cit. 16, 1, Jan73, 45-6) (proposes computer engineering and "amalgamation of theory and the practical approach").

KING, D. J. (1973) The Engineer - Programmer Relationship The Journal of the Institution of Engineers, Australia 45, 4-5, AprMay73, 9-12.

LITTLER, G. E. (1972) Engineering Technicians The Journal of the Institution of Engineers, Australia 44, 6, Jun72, 15-6,22.

LLOYD, BRIAN E. (1959) The Education of Professional Engineers in Australia Melbourne: The Association of Professional Engineers, Australia, 1968, 490pp.

McCLUSKEY, EDWARD J. (Chairman 1970) IEEE Computer Society Constitution & Bylaws IEEE Computer 3, 5, SepOct70, 33-8.

MINSKY, MARVIN (1970) Form and Content in Computer Science Journal of the ACM 17, 2, Apr70, 197-215.

NEWELL, ALLEN; PERLIS, ALAN J; SIMON, HERBERT (1967) What is Computer Science? Science 157, 1373-4 also in Finerman 1968 78-9.

PAPERT, SEYMOUR (1970) Teaching Children Thinking, in Scheepmaker & Zinn I/73-8, or Mathematics Teaching 58, Spr72, Association of Teachers of Mathematics.

PERLIS, ALAN J. (1968) Computer Science Is Neither Mathematics nor Electrical Engineering, in Finerman 69-77.

PICKERING, A. N. (1973) More Humanities for the Technologists? The Journal of the Institution of Engineers, Australia 45, 10-11, OctNov73, 13-4,22.

SACKMAN, HAROLD (1967) Computers, System Science, and Evolving Society New York: John Wiley & Sons, Inc., 638pp. (subtitled "The Challenge of Man-Machine Digital Systems", being largely based on experience with the SAGE system, an early and impressive example of data engineering).

SCHEEPMAKER, B; ZINN, KARL L. (Eds.1970) World Conference on Computer Education 1970 Amsterdam: International Federation for Information Processing, 136+472+280pp.

SCOTT, NORMAN R. (1973) Engineers are slipping IEEE Spectrum 10, 8, Aug73, 41-3.

SLAMECKA, VLADIMIR (1968) The Science and Engineering of Information, in Finerman 81-92.

SPOONER, MARGERET (1973) Study at UK polytechnics Computing 27Sep73, 18-9 (14 polytechnics covered, all B.Sc. or B.Sc.(Hons.) and 3 or 4 year course under such designations as Computer Science, Applied Computing, Statistics and Computing, and Computing Science), see also "Science course - or art" Computing 25Oct73, 15-7 (18 UK universities), and "Nineteen university ways to a 1st degree" Computing 22Nov73, 11-3.

STORR, E. D. (Ed.1972) Policy Review Committee: Extracts from Final Report The Journal of the Institution of Engineers, Australia 44, 4-5, Apr-May72, 26-9.

------ (1973a) Colleges of Engineers op.cit. 45, 6, Jun73, 2.

------ (Ed.1973b) The Minimum Duration of an Engineering Course op.cit., 20-2.

TEICHROW, DANIEL (Ed.1971) Education Related to the Use of Computers in Organizations Communications of the ACM 14, 9, Sep71, 573-88.

TOOTILL, GEOFFREY C. (Chairman 1968) IFIP-ICC Vocabulary of Information Processing Amsterdam: North - Holland Publishing Company, 1968, 208pp.

TOU, JULIUS T. (1970) Software Engineering - A New Profession Software Engineering Vol.I Ed. Julius T. Tou, New York: Academic Press, 1-6 ("these modern computers are so sophisticated and so general that when delivered they are ready for use on nothing, but they can be made useful on anything ... The users need the professional services of a specialist - the software engineer.").

WEGNER, PETER (1970) Three Computer Cultures: Computer Technology, Computer Mathematics, and Computer Science in Alt et al. 7-78 ("A science of information structure transformation may be developed that may well become as basic and important as mathematics or physics ..."(p.75)).

WILLINGHAM, D.J.S. (1973) Is The Programmer Dead ? Datafair 73 Conference Papers I, 9-12, British Computer Society.

ZEMANEK, HANS (1972) What is Informatics? Management Informatics 1, 2, Apr72, 43-52.